Available online at www.sciencedirect.com
SCIENCE DIRECT
ELSEVIER
Journal of Multivariate Analysis 96 (2005) 20–29
Journal of Multivariate Analysis
www.elsevier.com/locate/jmva

Correlation in L^p-spaces ☆

Aleksander Kowalski[a], Edmund Rudiuk[b,*]

[a] *Institute of Mathematics, Maria Curie-Sklodowska University, Pl. Marii Curie-Sklodowskiej 1, 20-031 Lublin, Poland*

[b] *Department of Physical and Applied Sciences, University College of Cape Breton, 1250 Grand Lake Road, Sydney, Nova Scotia, B1P 6L2 Canada*

Received 7 March 2003
Available online 18 November 2004

Abstract

An extension to an L^p-spaces, $p > 1$, of Pearson–Kolmogorov–Renyi correlation ratio is constructed. It is proved that correlation does not exceed $2^{\left|\frac{2}{p}-1\right|}$, and can be used as a measure of dependence of the random variable from a sigmafield.

AMS 1991 subject classification: 62H20; 60A99; 41A65

Keywords: Correlation; Measures of dependence; Metric projections; Banach spaces

1. Introduction

Correlation coefficient is a very popular and powerful measure of dependence between random variables. It is constructed inside L^2 theory of random variables [7] using methods from this theory. While dependence is more general property, that stretches beyond L^2, to the space of all random variables. Analyzing the results [3,4] we may conclude, that the classical correlation is nonadequate as a measure of dependence outside L^2-space. We need to construct a more universal measure of dependence (c.f. [13]), which is not restricted by the Hilbertian structure of L^2.

☆ Supported by Natural Sciences and Engineering Research Council of Canada.

* Corresponding author.
E-mail addresses: kowalskialekesander@yahoo.com (A. Kowalski), edmund_rudiuk@uccb.ca (E. Rudiuk).

0047-259X/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2004.09.006

Some of them are correlation and symmetric coefficient of covariation defined for symmetric α-stable (SαS) random variables. However, it cannot be extended over the space of SαS random variables.

The aim of this paper is to extend the correlation coefficient from Hilbert space L^2 of random variables with a finite second moment to the Banach space L^p, $p > 1$, of random variables with finite pth moments. The correlation must be consistent with orthogonality in the Banach spaces and must respect independency.

Leaving a Hilbert space we must forget about symmetricity of correlation, as orthogonality relation is not symmetric in the Banach spaces. Moreover, there exist a lot of different orthogonality concepts [9]. We will use traditional Birkhoff orthogonality definition, as it is fundamental in Banach spaces, very useful in measurement of dependence and estimation. It has at least three special properties distinguishing it from other:

- independent and centered random variables are orthogonal in the Birkhoff sense,
- it is intuitive in probability theory and natural in applications,
- it is connected to the metric projection operator.

We will also see, that the correlation coefficient, defined as a nonnegative number, is not bounded by one, but varying around one between zero and two. It is equal zero for independent random variables and one for completely dependent.

The classical correlation between indicators is often used as measure of dependence of random variables. We shall see that the L^p-correlation coefficient defines a measure of dependence which is bounded by $2^{\left|\frac{2}{p}-1\right|}$.

Let $H = L^2(\mathcal{F})$ denote a Hilbert space of all real random variables defined on the complete probability space $(\Omega, \mathcal{F}, P)$ with finite second moments. The random variables are treated as elements (vectors) of a linear space equipped with the Hilbertian structure. The correlation coefficient between random variables is introduced as a cosine of an angle between vectors and is defined in two ways: using a scalar product or using an orthogonal projection operator.

The first approach is very popular in literature (c.f. [13]) while the second one was introduced by K. Pearson and A. N. Kolmogorov, and studied by A. Rényi in the context of a maximal correlation [7,8]. We recall these definition.

Let $X \in H,\ Y \in H$ be a pair of random variables with finite second moments. H^Y stands for the linear space spanned by the random variable Y and $P(\cdot|H^Y)$ denote the orthogonal projection operator onto that linear space.

The correlation coefficient $\rho(X, Y)$ between random variables X and Y is a real number

$$\rho(X, Y) = \frac{cov(X, Y)}{\sigma(X)\sigma(Y)}, \tag{1}$$

where $cov(X, Y) = E[(X - EX)(Y - EY)]$ is the scalar product between centered random variables, while $\sigma(X) = \sqrt{E[(X - EX)^2]}$ and $\sigma(Y) = \sqrt{E[(Y - EY)^2]}$ are norms of centered random variables.

Let now $\mathcal{G} \subset \mathcal{F}$ be a sub-σ-field of $\mathcal{F}$. Then the Hilbert space $H_1 = L^2(\mathcal{G})$ is a subspace of H, and the conditional expectation $E_{\mathcal{G}}$ is an orthogonal projection operator onto that space. This yields the definition of the correlation coefficient as a correlation ratio [7,8]

$\rho(X, \mathcal{G})$ between the random variable X and the space $L^2(\mathcal{G})$:

$$\rho(X, \mathcal{G}) = \frac{\sigma(E_{\mathcal{G}}X)}{\sigma(X)}. \tag{2}$$

In this approach the cosine between the random variable X and the linear space H_1 is computed from a right triangle as a ratio of the length of projection to the length of the projected vector. It is not difficult to check that the correlation coefficient is invariant with respect to nonrandom translation and rescaling. Namely, we have: $\rho(aX+b, \mathcal{G}) = \rho(X, \mathcal{G})$ for $a \neq 0, a, b \in R$.

If the σ-field $\mathcal{G}$ is generated by the random variable Y then $E_{\mathcal{G}}X = E[X|Y]$, and the last defined correlation may be considered as a measure of dependence of the random variable X from Y (c.f. [7]). Putting $X = 1_A,\ Y = 1_B,\ A, B \in \mathcal{F}$ the correlation coefficient may be treated as a measure of dependence between random events A and B.

However, as it is explicitly shown in the paper [4], the correlation coefficient is extremely asymmetric.

Let now $L^p(\mathcal{F}),\ p > 1$ denote the Banach space of all real random variables, with finite pth moments, and defined on a complete probability space $(\Omega, \mathcal{F}, P)$. We shall consider extension of the definition of the correlation coefficient from a Hilbert space L^2 to the Banach space L^p. Since a scalar product in Banach spaces is not defined, then the direct extension of definition (1) is impossible.

Working in the L^2-space we see that family of Gaussian random variables play a fundamental role. The families of random variables with symmetric α-stable distributions (SαS) are their generalization up to some larger spaces. There are many intuitive properties of Gaussian distributions that are shared by the symmetric α-stable distributions. The family of SαS random variables forms a linear space, similarly to the Gaussian case.

The correlation coefficient between random variables, without second moment assumption, has been introduced in [6] for random variables with a symmetric α-stable joint distribution (SαS random variables). It was studied and evolved in paper [2].

We shall quote some results following this paper.The pair (X, Y) of random variables has a SαS, $\alpha \in (0, 2\rangle$, distribution ([2, Definition 1]) if its characteristic function φ is given by

$$\varphi_{(X,Y)}(s, t) = \int_S \exp(-|sx + ty|^{\alpha})\, d\mu_S(x, y),$$

where μ_S is a symmetric measure (called a spectral measure) on the borelians of the unit circle $S = \{(x, y) \in R^2 : x^2 + y^2 = 1\}$.

Extending definition of covariance, the covariation between SαS random variables is introduced in the following manner:

First, the covariation between SαS random variables X and Y is equal to

$$[X, Y]_{\alpha} = \int_S x|y|^{\alpha-1} sgn(y)\, d\mu_S(x, y), \tag{3}$$

where μ_S is the spectral measure on the circle S.

Then the coefficient $\{X, Y\}_{\alpha}$ of covariation of X on Y is defined as

$$\{X, Y\}_{\alpha} = \frac{[X, Y]_{\alpha}}{[Y, Y]_{\alpha}}. \tag{4}$$

Properties of the above quantities were studied in [10]. The definition has several disadvantages which were described in [2]. As a proper modification of (4) they introduced a new measure of covariation for stable random variables.

A symmetric coefficient of variation between SαS random variables X and Y is equal to

$$Corr_\alpha(X, Y) = \frac{[X, Y]_\alpha [Y, X]_\alpha}{[X, X]_\alpha [Y, Y]_\alpha}. \tag{5}$$

It was proved [2] that $Corr_\alpha(X, Y)$ is symmetric, its module not exceed one and equals zero if X, Y are independent. Moreover, for Gaussian random variables a symmetric coefficient of variation equals the square of classical correlation coefficient: $Corr_\alpha(X, Y) = \rho^2(X, Y)$.

2. Orthogonality and deviation in L^p-spaces

The correlation in L^2-space was defined as the cosine of angle between random variables, or between random variable and a linear space. The scalar product or the ratio of lengths were used in this definition.

For symmetric and α-stable random variables, which may be beyond L^2-space, the form $[X, Y]_\alpha$ is used as a some asymmetric substitution of a scalar product and $[Y, Y]_\alpha$ as a substitution of a norm in α-power (cf. [2]). Then the coefficient of covariation, as an extension of the correlation is defined using idea coming from the classical formula (1).

In our construction we shall walk following the idea of a correlations ratio (2) which comes from Pearson, Kolmogorov and Rényi [7].

We start with the orthogonality concepts, which in Banach spaces is not that obvious that it is in Hilbert spaces. In the L^p-space we have several different definitions of orthogonality [11,9]. We shall use the Birkhoff definition as fundamental and the most useful in the probability theory framework.

Let $L^p(\mathcal{F})$, $p > 1$, be a Banach space of random variables defined on $(\Omega, \mathcal{F}, P)$ with the finite norm: $\|X\| = \sqrt[p]{E|X|^p}$ (we do not distinguish random variables coinciding a.s.). For a given sub σ-field $\mathcal{G} \subset \mathcal{F}$ let $L^p(\mathcal{G})$ denote the linear space of all $\mathcal{G}$-measurable random variables with finite pth moments. The set of all real numbers R is identified with the set of all constant random variables. Hence $R = L^p(\sigma(\emptyset, \Omega))$ and inclusions $R \subset L^p(\mathcal{G}) \subset L^p(\mathcal{F})$ of linear closed spaces hold true.

Definition 1. The random variable X is orthogonal to Y, $X \perp Y$, $X, Y \in L^p(\mathcal{F})$ if for every $a \in R$:

$$\|X\| \leqslant \|X + aY\|. \tag{6}$$

The random variable $X \in L^p(\mathcal{F})$ is orthogonal to the space $L^p(\mathcal{G})$, $\mathcal{G} \subset \mathcal{F}$: $X \perp L^p(\mathcal{G})$ if $X \perp Y$ for every $Y \in L^p(\mathcal{G})$, i.e. for every $Y \in L^p(\mathcal{G})$

$$\|X\| \leqslant \|X + Y\|. \tag{7}$$

The distance of a random variable $X \in L^p(\mathcal{F})$ from the space $L^p(\mathcal{G})$ is a real number

$$e_{\mathcal{G}}(X) = \inf_{Y \in L^p(\mathcal{G})} \|X - Y\| \tag{8}$$

We call the deviation of a random variable $X \in L^p(\mathcal{F})$ its distance from *R*:

$$e(X) = e_{\sigma(\phi,\Omega)}(X) = \inf_{\mu \in R} \|X - \mu\|. \tag{9}$$

In the special case, when $p = 2$, the deviation is the classical standard deviation of a random variable. The following theorem collects fundamental properties of the distance of $X \in L^p(\mathcal{F})$ to the linear space of random variables $L^p(\mathcal{G})$. The proof can be found in the Singers monograph [11, Theorem 6.5 and Corollary 3.5].

Theorem 1. *Let $X \in L^p(\mathcal{F})$, $p > 1$, and $\mathcal{G} \subset \mathcal{F}$, then the distance $e_{\mathcal{G}}(X)$ has the following properties*:

e1.

$$\begin{aligned} 0 &\leqslant e_{\mathcal{G}}(X) < \infty; \\ e_{\mathcal{G}}(X) &= 0 \Leftrightarrow X \in L^p(\mathcal{G}); \end{aligned} \tag{10}$$

e2.

$$\begin{aligned} e_{\mathcal{G}}(X) &\leqslant \|X - Y\|, \ for\ every\ Y \in L^p(\mathcal{G}); \\ e_{\mathcal{G}}(X) &\leqslant \|X\|, \quad e_{\mathcal{G}}(X) = \|X\|\ if\ X \perp \mathcal{G}; \end{aligned} \tag{11}$$

e5.

$$e_{\mathcal{G}}(X + Y) = e_{\mathcal{G}}(X), \quad X \in L^p(\mathcal{F})\ and\ Y \in L^p(\mathcal{G}); \tag{12}$$

$$e_{\mathcal{G}}(aX + b) = |\alpha| e_{\mathcal{G}}(X)\ for\ a, b \in R; \tag{13}$$

e7.

$$|e_{\mathcal{G}}(X) - e_{\mathcal{G}}(Y)| \leqslant \|X - Y\|, \quad X, Y \in L^p(\mathcal{F}) \tag{14}$$

and functional $e_{\mathcal{G}}$ is continuous.

e8. *the infimum* (8) *is attained at a single element of the space $L^p(\mathcal{G})$ i.e.the set*

$$\left\{ \hat{X} \in L^p(\mathcal{G}) : \|X - \hat{X}\| = \inf_{Y \in L^p(\mathcal{G})} \|X - Y\| \right\}$$

has one element.

The last point of the theorem yields to consider (cf. [1]) the metric projection of a random variable *X* into $L^p(\mathcal{G})$ as a $\mathcal{G}$-measurable random variable for which infimum (8) is attained. To achieve analogies with an L^2-space the map assigning $\hat{X}$ to X we call the conditional L^p-expectation. The conditional L^p-expectation is direct generalisation of the classical conditional expectation which is the conditional L^2-expectation. We introduce the following definition:

Definition 2. The conditional L^p-expectation, $p > 1$, is the operator $\mathbf{E}_{\mathcal{G}} : L^p(\mathcal{F}) \to L^p(\mathcal{G})$

$$\mathbf{E}_{\mathcal{G}}[X] = \arg \inf_{Y \in L^p(\mathcal{G})} \|X - Y\|. \tag{15}$$

If $\mathcal{G} = \sigma(\phi, \Omega)$ is a trivial σ-field, then the operator $\mathbf{E} = \mathbf{E}_{\sigma(\phi,\Omega)}$ will be called the L^p-expectation.

We note that

$$e_{\mathcal{G}}(X) = \|X - \mathbf{E}_{\mathcal{G}}[X]\| \tag{16}$$

and

$$\|X - \mathbf{E}_{\mathcal{G}}[X]\| \leqslant \|X - Y\|, \text{ for every } Y \in L^p(\mathcal{G}). \tag{17}$$

Now, considering Eq. (7), we get the following characterization of the conditional L^p-expectation (cf. [11, Lemma 1.14]).

Corollary 1. $\mathbf{E}_{\mathcal{G}}X = \hat{X}$ *iff* $\hat{X} \in L^p(\mathcal{G})$ *and* $X - \hat{X} \perp L^p(\mathcal{G})$.

The following theorem characterizes the orthogonality in L^p-spaces with $p > 1$. (cf. [11, Theorem 1.11]).

Theorem 2. *The random variable* $X \in L^p(\mathcal{F})$ *is orthogonal to* $Y \in L^p(\mathcal{F})$ *iff*

$$E[|X|^{p-1} sgn(X)Y] = 0. \tag{18}$$

$X \in L^p(\mathcal{F})$ *is orthogonal to* $L^p(\mathcal{G})$ *iff*

$$E[|X|^{p-1} sgn(X)1_A] = 0 \tag{19}$$

for every event $A \in \mathcal{G}$.

The proof of theorem can be found in [11, Theorem 1.11].
We see that $\mathbf{E}_{\mathcal{G}}[X] = \hat{X}$ iff $\hat{X} \in L^p(\mathcal{G})$ and for every $A \in \mathcal{G}$ we have

$$E[|X - \hat{X}|^{p-1} sgn(X - \hat{X})1_A] = 0. \tag{20}$$

To compute an L^p-expectation we will use a special case of the above equation:

$$\mathbf{E}[X] = \mu \text{ iff } E[|X - \mu|^{p-1} sgn(X - \mu)] = 0. \tag{21}$$

We prove the following proposition.

Proposition 1. Let $X \in L^p(\mathcal{F})$ *be a random variable independent from* σ-field $\mathcal{G}$. *Then* $X - \mathbf{E}[X] \perp \mathcal{G}$.

We note that above theorem states, that under independence assumption, we have: $\|X - \mathbf{E}[X]\| \leqslant \|X - Y\|$ for every $Y \in L^p(\mathcal{G})$.

Proof. By the assumption we have

$$\begin{aligned} &E[|X - \mathbf{E}[X]|^{p-1} sgn(X - \mathbf{E}[X])1_A] \\ &\quad = E[|X - \mathbf{E}[X]|^{p-1} sgn(X - \mathbf{E}[X])]P(A) = 0. \end{aligned}$$

using Corollary 1 and (20), since $X - \mathbf{E}[X] \perp R$. $\square$

Theorem 3. *Let $L^p(\mathcal{F})$, $1 < p < \infty$, be a Banach space of random variables and let $L^p(\mathcal{G})$, $\mathcal{G} \subset \mathcal{F}$, stands for its subspace. Then for $X \in L^p(\mathcal{F})$ we have*:

E1. $\mathbf{E}_{\mathcal{G}}[X] = X, \quad X \in L^p(\mathcal{G})$;

E2.

$$\|\mathbf{E}_{\mathcal{G}}[X]\| \leqslant 2^{|\frac{2}{p}-1|}\|X\|; \tag{22}$$

E3. *if* $\max\{\|X\|, \|Y\|\} \leqslant r$ *then*

$$\|\mathbf{E}_{\mathcal{G}}[X] - \mathbf{E}_{\mathcal{G}}[Y]\| \leqslant k\|X - Y\|, \tag{23}$$

where $k = dr^{1-q}$, $\quad d \leqslant 2 + c^{-q}$, $\quad c = 4^{-1}p(p-1)1_{[1<p<2]} + 2^{1-p}1_{[p \geqslant 2]}$, $q = \max\{2, p\}$;

E4.

$$\mathbf{E}_{\mathcal{G}}[X + Y] = \mathbf{E}_{\mathcal{G}}[X] + Y, \; for X \in L^p(\mathcal{F}), \;\; Y \in L^p(\mathcal{G}), \tag{24}$$

$$\mathbf{E}_{\mathcal{G}}[aX + b] = a\mathbf{E}_{\mathcal{G}}[X] + b, \;\; a, b \in R; \tag{25}$$

E5.

$$\mathbf{E}_{\mathcal{G}}[\mathbf{E}_{\mathcal{G}_1}[X]] = \mathbf{E}_{\mathcal{G}_1}[X] \; if \; \mathcal{G}_1 \subset \mathcal{G}; \tag{26}$$

E6. *if a random variable X is bounded with probability* 1, *and*: $m = \inf_{\omega\in\Omega} ess X(\omega) \leqslant X(\omega) \leqslant \sup_{\omega\in\Omega} ess X(\omega) = M, n, M \in R$, *then*

$$m \leqslant \mathbf{E}_{\mathcal{G}}[X] \leqslant M. \tag{27}$$

Proof. The proof of the properties E1, E4 and E5 can be found in [1,11], E2 is proved in [5], and E3 in [12].

To prove E6 note that if $\hat{X} < m$ on a some event, then the random variable $\check{X} = \max\{\hat{X}, m\} \in L^p(\mathcal{G})$ is closer to X than is $\hat{X}$ which contradicts that $\hat{X}$ is closest to X element of $L^p(\mathcal{G})$. Then it must be $\hat{X} \geqslant m$ on that event. Analogously we prove that $\hat{X} \leqslant M$. □

We note that in general $\mathbf{E}_{\mathcal{G}}$ is not a linear operator and equality (26) does not hold true if we change the order of projections in the left side of this equality.

3. Correlation in L^p-spaces

Let $L^p(\mathcal{F})$, $p > 1$ be a Banach space of random variables defined on $(\Omega, \mathcal{F}, P)$ and for a given sub σ-field $\mathcal{G} \subset \mathcal{F}$ let $L^p(\mathcal{G})$ denote a linear subspace of all $\mathcal{G}$ -measurable random variables.

Following the classical definition of the correlation coefficient, as a correlation ratio, we may extend this concept to the space $L^p(\mathcal{F})$.

Definition 3. L^p-correlation coefficient between nondegenerate random variable $X \in L^p(\mathcal{F})$ and a space $L^p(\mathcal{G})$ is the ratio:

$$\rho(X, L^p(\mathcal{G})) = \frac{e(\hat{X})}{e(X)}, \tag{28}$$

where $\hat{X} = \mathbf{E}_{\mathcal{G}}X$ is a conditional L^p-expectation of X, and $e(X) = \|X - \mathbf{E}[X]\|$, $e(\hat{X}) = \|\hat{X} - \mathbf{E}\hat{X}\|$ are L^p-deviations of random variables X and $\hat{X}$, respectively.

The L^p-correlation coefficient between the random variable X and the random variable Y is defined as an L^p-correlation between X and the space $L^p(\mathcal{G})$ where $\mathcal{G}$ is a σ-field generated by Y:

$$\rho(X, Y) = \frac{e(\mathbf{E}[X|Y])}{e(X)}. \tag{29}$$

The correlation coefficient is the ratio of deviations: the conditional expectation $\mathbf{E}_{\mathcal{G}}X$ of X, and the random variable X.

Using (9), correlation (28) may be expressed as

$$\rho(X, L^p(\mathcal{G})) = \frac{\inf_{\mu\in R} \|\mathbf{E}_{\mathcal{G}}X - \mu\|}{\inf_{\mu\in R} \|X - \mu\|}.$$

Some properties of the L^p-correlation coefficient follow from Theorems 1 and 3. We collect them in the following theorem:

Theorem 4. *Let $\mathcal{G} \subset \mathcal{F}$ be a sub-σ-field of $\mathcal{F}$. Then for every random variable $X \in L^p(\mathcal{F})$ we have*:

C1. $\rho(X, L^p(\mathcal{G})) \geqslant 0$.

C2. $\rho(X, L^p(\mathcal{G})) = 0 \Longleftrightarrow X - \mathbf{E}X \perp L^p(\mathcal{G})$.

C3. $\rho(X, L^p(\mathcal{G})) = 0$ *if the random variable X is independent from the σ-field $\mathcal{G}$.*

C4. *If $X \in L^p(\mathcal{G})$ then $\rho(X, L^p(\mathcal{G})) = 1$.*

C5. *For every $a \neq 0$, $b \in R$*:

$$\rho(aX + b, L^p(\mathcal{G})) = \rho(X, L^p(\mathcal{G})). \tag{30}$$

C6.

$$\rho(X, L^p(\mathcal{G})) \leqslant 2^{|\frac{2}{p}-1|}. \tag{31}$$

C7. $\rho(X, L^p(\mathcal{G}))$ *is continuous functional of X for $e(X) > 0$, i.e.*

$$\rho(X_n, L^p(\mathcal{G})) \to \rho(X, L^p(\mathcal{G})) \mathit{if}\, X_n \to X \mathit{\ in\ } L^p \text{-}\mathit{norm\ and\ } e(X) > 0. \tag{32}$$

Proof. C1. $\rho(X, L^p(\mathcal{G}))$ is a ratio of non-negative numbers.

C2. The random variable $X - \mathbf{E}[X]$ is orthogonal in Birkhoff sense to the space $L^p(\mathcal{G})$ iff $\mathbf{E}_{\mathcal{G}}[X - \mathbf{E}[X]] = 0$ (7). Using now (25), (17) we obtain that $0 = \|\mathbf{E}_{\mathcal{G}}[X - \mathbf{E}X]\| = \|\mathbf{E}_{\mathcal{G}}[X] - \mathbf{E}X\| \geqslant \|\mathbf{E}_{\mathcal{G}}[X] - \mathbf{E}[\mathbf{E}_{\mathcal{G}}[X]]\| \geqslant 0$. It is equivalent that $e(\hat{X}) = 0$ i.e. $\rho(X, L^p(\mathcal{G})) = 0$.

C3. The proof follows directly from just proved property C2 and Proposition 1. Suppose that X is independent from $\mathcal{G}$. Then $\mathbf{E}_{\mathcal{G}}[X] = \mathbf{E}[X]$ and $e(\hat{X}) = 0$.

C4. If $X \in L^p(\mathcal{G})$ then, by the property E1 of the conditional expectation, we have $\hat{X} = \mathbf{E}_{\mathcal{G}}[X] = X$, and $e(\hat{X}) = e(X)$ which proves C4.

C5. Put $Y = aX + b$. Then, by (13) we have $e(Y) = |a|e(X)$. Next, by (25)

$$\widehat{Y} = \mathbf{E}_{\mathcal{G}}[Y] = a\mathbf{E}_{\mathcal{G}}[X] + b = a\hat{X} + b$$

and similarly as for Y, we get $e(\widehat{Y}) = |a|e(\hat{X})$. Hence

$$\rho(Y, L^p(\mathcal{G})) = \frac{e(\widehat{Y})}{e(Y)} = \frac{|a|e(\hat{X})}{|a|e(X)} = \rho(X, L^p(\mathcal{G})),$$

and property (30) is proved.

C6. We just proved that the correlation coefficient is invariant with respect to the non-random rescaling and translations. So, we may assume that $\mathbf{E}[X] = 0$. Taking into account (11) and (22) we have:

$$e(\hat{X}) \leqslant \|\hat{X}\| \leqslant 2^{|\frac{2}{p}-1|}\|X\| = 2^{|\frac{2}{p}-1|}e(X),$$

which implies (31) and proves C6.

C7. Suppose that $X_n \to X$ in L^p-norm and $e(X) > 0$. Then using property E3 we get $\mathbf{E}_{\mathcal{G}}[X_n] \to \mathbf{E}_{\mathcal{G}}[X]$ in L^p -norm. Then by property (14) $e(\mathbf{E}_{\mathcal{G}}[X_n]) \to e(\mathbf{E}_{\mathcal{G}}[X])$, and $e(X_n) \to e(X)$, $e(X_n) > 0$ for sufficiently large n. Hence

$$\rho(X_n, L^p(\mathcal{G})) = \frac{e(\mathbf{E}_{\mathcal{G}}[X_n])}{e(X_n)} \to \frac{e(\mathbf{E}_{\mathcal{G}}[X])}{e(X)} = \rho(X, L^p(\mathcal{G})),$$

which proves C7. $\square$

4. Conclusions

A correlation coefficient is a one of several dependence measures between random variables and between random events. The fundament of probability theory and mathematical statistics is considering random variables as measurable functions. Measurable functions form functional spaces. Canonical functional spaces are L^p spaces. Independence property of random variables is invariant property through all spaces L^p (as long as variables belongs to these spaces). Constructing measures of dependence we should assume that absolute value of the measure may vary from zero to two.

Following ideas from A. Rényi papers [7,8] the maximal coefficient of correlation may be introduced and studied. It will also vary between zero and two and equals zero iff the random variables are independent.

We note that the classical correlation and maximal correlation coefficient are defined and constructed inside the L^2-theory of random variables using methods which come from the theory of Hilbert spaces.

The maximum correlation coefficient, the measure of dependence between two measurable functions, is defined using average, scalar product and the L^2-norm of functions. It is very special measure of dependence, strictly and uniquely associated with the L^2-norm (c.f. [3]). If the distance between random variables (i.e. measurable functions) is measured using the more general L^p-norm, then the measure of dependence must be redefined using this norm and average instead of the L^2-norm and arithmetic average.

The section of probability theory in which the classical conditional expectation, correlation and maximal correlation is considered was entitled as "The Hilbert Space of Random Variables" or as "L^2-theory", while the section which contains the concepts of the conditional L^p-expectation and L^p-correlation should be entitled as "The Banach Space of Random Variables" or as an "L^p-theory". The relation between these sections are similar to relations between theory of Hilbert spaces and theory of Banach spaces.

Acknowledgments

The authors are very grateful to the referee whose comments helped to improve the paper.

References

[1] T. Ando, I. Amemiya, Almost everywhere convergence of prediction sequences in L^p, $p \in (1, \infty)$, Z. Wahrscheinlichketstheorie und Verw. Gebiete 4 (1965) 113–120.
[2] L. dEstampes, B. Garel, D. Tjøstheim, On a new measure of covariation for stable random variables, 2003, Preprint submitted to Elsevier Preprint.
[3] A. Kagan, L.A. Shepp, Why the variance?, Statist. Probab. Lett., in press.
[4] A. Kagan, H. Furstenberg, Linear regression and second moments. Statist. Probab. Lett. 62 (2003), in press.
[5] F. Mazzone, On the norm of the metric projections, J. Approx. Theory 97 (1999) 214–219.
[6] G. Miller, Properties of certain symmetric stable distributions, J. Multivariate Anal. 8 (3) (1978) 346–360.
[7] A. Rényi, New version of the probabilistic generalization of the large sieve, Acta Math. Acad. Sci. Hungar. 10 (1959) 217–226.
[8] A. Rényi, On measures of dependence, Acta Math. Acad. Sci. Hungar. 10 (1959) 441–451.
[9] F.B. Saidi, An extension of the notion of orthogonality to Banach spaces, J. Math. Anal. Appl. 267 (1) (2002) 29–47.
[10] G. Samorodnitsky, M.S. Taqqu, Stable non-Gaussian random processes, Stochastic models with infinite variance, Stochastic Modeling, Chapman & Hall, New York, 1994.
[11] I. Singer, Best Approximation in Normed Linear Spaces by Elements of Linear Subspaces, Springer, Berlin, 1970.
[12] R. Smarzewski, Strongly unique best approximation in Banach spaces, J. Approx. Theory 47 (3) (1986) 184–194.
[13] D. Tjøstheim, Measures of dependence and tests of independence, Statistics 28 (3) (1996) 249–284.